AD-A160 030   THE INDEXING PROPERTIES OF AN ANCILLARY STATISTIC(U)   1/1
STANFORD UNIV CA DEPT OF PHYSICS   A Y KUK 28 AUG 85
TR-363 N00014-76-C-0475
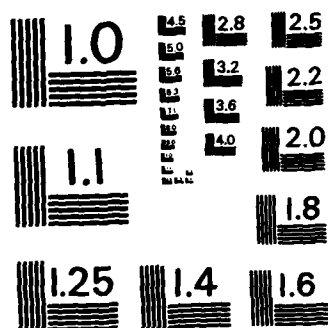
UNCLASSIFIED   F/G 12/1   NL

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

THE INDEXING PROPERTIES OF AN ANCILLARY STATISTIC

BY

ANTHONY Y.C. KUK

TECHNICAL REPORT NO. 363

AUGUST 28, 1985

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

DTIC
ELECTE
OCT 10 1985
B

85 10 8 077

THE INDEXING PROPERTIES OF AN ANCILLARY STATISTIC

BY

ANTHONY Y.C. KUK

TECHNICAL REPORT NO. 363
AUGUST 28, 1985

DEPARTMENT OF STATISTICS
STANFORD   UNIVERSITY
STANFORD, CALIFORNIA

DTIC
ELECTE
OCT 10 1985
B

The Indexing Properties Of An Ancillary Statistic

Anthony Y.C. Kuk

1. INTRODUCTION

Approximate ancillarity has recently been the subject of substantial work (Efron & Hinkley 1978; Cox 1980; Hinkley 1980; Barndorff-Nielsen 1980; Amari 1982b). While this topic is undoubtedly important, we believe that the concept of exact ancillarity has not yet been fully explored. The usual presentation consists mainly of examples of ancillary statistics, on the basis of which certain statements are made. These statements, though often appealing, are hard to make precise. The purpose of our study is to clarify some of the properties of an ancillary statistic.

In §2, we reexamine some of the examples and statements in a way relevant to our subsequent discussion. In §3, we introduce transformation models and point out that most of the known examples of ancillary statistics fall within this category. We also give an interpretation of the statement: "Two samples with the same ancillary statistic value contain equal amount of information." In §4, we discuss the role of an ancillary statistic as a precision index. For transformation models, and also for exponential models, for which ancillary statistics also exist, we find that the variance of the conditional Fisher information is proportional to the square of the statistical curvature γ_θ (Efron, 1975). Thus the magnitude of γ_θ is a measure of the effect of conditioning. Moreover, the constancy of γ_θ as a function of θ appears related to the concept of exact precision index (Buehler, 1982). We conclude in §5 with some miscellaneous remarks.

2. EXAMPLES

The standard examples used to introduce the concept of ancillary statistics and the conditionality principle are those of random sample size, of two measuring instruments and the mixture problem (Cox & Hinkley 1974, pp.32,38). These examples are intended to suggest that the observed value of the ancillary statistic describes the part of the total sample space relevant to the problem at hand, and that inference about the parameter should be conditioned on that value. In the standard examples, the fact that some other sample size, some other instrument or some other distribution might have been used, but actually was not, is irrelevant. While these examples may seem artificial, similar situations arise more subtly in other problems of statistical inference.

Example 1. Fisher's normal circle. Efron (1978), Efron & Hinkley (1978). Let $X_1, \ldots, X_n$ be independent observations from a bivariate normal distribution with mean vector $(\rho\cos\theta, \rho\sin\theta)$ lying on a circle of given radius ρ and with identity covariance matrix; then $X = \Sigma X_i$ is sufficient. If X has polar coordinates $(\hat{\theta}, r)$, then $\hat{\theta}$ is the maximum likelihood estimate of θ and r is ancillary. It can be shown that $E_\theta\{(\hat{\theta}-\theta)^2\}$ and $E_\theta\{(\hat{\theta}-\theta)^2|r\}$ are both independent of θ and that the latter is a decreasing function of r, so that the accuracy of $\hat{\theta}$ improves as r increases.

Example 2. Normal mean with known coefficient of variation. Hinkley (1977). Let $X_1, \ldots, X_n$ be independently $N(\mu, b^2\mu^2)$ where b is a known constant and $\mu > 0$; then $(\Sigma X_i, \Sigma X_i^2)$ is sufficient and $C = (\Sigma X_i^2)^{\frac{1}{2}}/\Sigma X_i$ is ancillary. Hinkley (1977) shows that the unconditional and conditional Fisher information about $\theta = \log\mu$ are both free of θ and that the conditional information is a decreasing function of c.

Examples 1 and 2 suggest that even though an ancillary statistic by itself carries no information about θ, it is of value when used in conjunction with some other statistic. To be more precise, suppose that the minimal sufficient statistic can be written as $S = (T,C)$ where C is the ancillary part and T is the so-called information carrying part. Although C contains no information about θ, it constitutes part of the minimal sufficient statistic. A good way to utilise C is to carry out conditional inference, and so the role of C as a precision index becomes relevant. The expression of S in the form (T,C) can happen only when a model is not complete because by a theorem of Basu (1955, 1958), if S is boundedly complete, then S cannot contain any ancillary component. Related to this discussion is a result of Lehmann (1981) which can be summarized roughly by saying that the various forms of completeness of a sufficient statistic S characterize the success of S in separating the informative part of the data from that part which by itself carries little or no information.

In other examples, it is obvious that some data values are more informative than others, as in the example of random sample size (Efron, 1978). Another example is the following.

Example 3. Location parameter of a uniform distribution. Let $X_1, \ldots, X_n$ be independently $U(\theta - \tfrac{1}{2}, \theta + \tfrac{1}{2})$; then the minimal sufficient statistic is the pair of extreme order statistics $(X_{(1)}, X_{(n)})$ and an ancillary statistic is the sample range $C = X_{(n)} - X_{(1)}$. If the observed c is close to 1, we can almost pinpoint θ whereas if c is close to zero, the sample is relatively uninformative.

Fisher (1935, p.48) wrote: "Ancillary statistics are only useful when different samples of the same size can supply different amounts of

information, and serve to distinguish those which supply more from those which supply less." Thus a useful ancillary statistic divides the sample space into equally informative subsets. Since some samples are more informative than others, we should not average over the whole sample space to obtain an inference, but only over those samples that contain the same amount of information. This leads again to the conditionality principle. It is, however, very hard to make precise the above idea, as remarked by Efron (1978): "So far, it has proved impossible to codify this statement in a satisfactory way." In the next section, we will give one interpretation of this statement within the context of transformation models.

Fisher's statement also leads to the idea that the better an ancillary statistic is in distinguishing the more informative samples from those that are less so, the more useful it is as a conditioning variable. An implementation of this criterion for choosing an ancillary has been provided by Cox (1971), who compares different ancillaries in terms of $\text{var}\{I_\theta(C)\}$, the variance of the conditional Fisher information. Properties of this criterion are studied by Becker & Gordon (1983). In §4, we shall see that for both transformation and exponential models, $\text{var}\{I_\theta(C)\}$ is proportional to γ_θ^2, the square of the statistical curvature.

3. TRANSFORMATION MODELS

3.1 *Ancillarity of maximal invariant statistic*

Suppose that we have a problem invariant under a group of transformations G, and that preliminary reduction by sufficiency has already taken place so that G acts on the space of sufficient statistics S with maximal invariant statistic C(S). Let G^* be the group of induced transformations on the parameter space and $\nu(\theta)$ the corresponding maximal invariant function. If $\nu(\theta)$ is a constant function, then C(S) is ancillary. Mathematically, if $\nu(\theta)$ is a constant function, the model is equivalent to a structural model (Fraser 1968) even though the emphasis is quite different. Barndorff-Nielsen (1980) calls such models trans-formation models.

3.2 *Examples*

Most of the examples of ancillary statistics actually fall under this category. Suprisingly, we have not been able to find any mention of this fact in the literature; instead, we find examples scattered around and treated separately. It seems worthwhile, therefore, to look at some of those examples from our present point of view, starting with the list given by Buehler (1982). In addition to our Examples 1, 2 and 3, we find the following.

Example 4. Sprott (1961). Let $X_1, \ldots, X_n$ be independently $Exp(ae^{k\theta})$, where a and k are known constants. The X's and Y's are independent. This problem remains invariant under transformations $X'_i = X_i + b$, $Y'_j = e^{bk}Y_j$ or equivalently $\log Y'_j = \log Y_j + bk$ ($i = 1,\ldots,n$; $j = 1,\ldots,m$). The minimal sufficient statistic is $(S_1,S_2) = (\Sigma X_i, \Sigma Y_j)$, the maximal invariant statistic is $C(S_1,S_2) = S_1/n - (\log S_2)/k$ and $\nu(\theta)$

is a constant function. Hence C is ancillary.

Example 5. Fisher's gamma hyperbola. Let $X \sim \text{Exp}(\theta)$, $Y \sim \text{Exp}(1/\theta)$ independent of X. Observe n pairs (X_1,Y_1), ..., (X_n,Y_n). This problem remains invariant under transformations $X_i' = bX_i$, $Y_i' = Y_i/b$, $b > 0$ $(i = 1,...,n)$. The sufficient statistic is $(S_1,S_2) = (\Sigma X_i, \Sigma Y_i)$, the maximal invariant statistic is $C(S_1,S_2) = S_1 S_2$ and $\nu(\theta)$ is a constant function. Hence C is ancillary.

The following example seems to be new.

Example 6. A special case of a correlated bivariate gamma density given by Barndorff-Nielsen (1980) is

$$f(x,y;\alpha,\theta) = (\alpha\theta - 1)I_0(2\sqrt{xy})\exp(-\alpha x - \theta y) \quad (x > 0, y > 0). \qquad (3.1)$$

The parameter space is $\alpha > 0$, $\theta > 0$, $\alpha\theta > 1$ and $I_0(\sqrt{u})$ is the Bessel function $\Sigma u^j/(j!)^2$. If α, θ are restricted by $\alpha\theta = a$ where $a > 1$, the bivariate density becomes

$$f(x,y;\theta) = (a - 1)I_0(2\sqrt{xy})\exp(-ax/\theta - \theta y) \quad (x > 0, y > 0). \qquad (3.2)$$

If we have n independent pairs of bivariate observations (X_1,Y_1), ..., (X_n,Y_n) from density (3.2), then the problem remains invariant under transformation $Y_i' = bY_i$, $X_i' = X_i/b$, $b > 0$ $(i = 1,...,n)$. The sufficient statistic is $(S_1,S_2) = (\Sigma X_i, \Sigma Y_i)$, the maximal invariant statistic is $C(S_1,S_2) = S_1 S_2$ and $\nu(\theta)$ is a constant function. So C is ancillary.

The mathematical equivalence of transformation models and structural models can also be used to construct new examples of ancillary statistics. There is, however, a slight complication since the structural approach is not concerned with sufficiency reduction. For example, if X_1, ..., X_n are independently $N(\mu,1)$ and we do not reduce by sufficiency,

$C = (X_{(2)} - X_{(1)}, \ldots, X_{(n)} - X_{(1)})$ is ancillary but if we reduce X_1, $\ldots$, X_n to the sufficient statistic $S = \Sigma X_i$, an ancillary statistic no longer exists. It follows from our discussion in §2 that we should look for a structural model which is not complete.

Example 7. Progression model. The progression model is a multivariate structural model introduced by Fraser (1968, p.139). We consider the simplest case with dimension two, the group G indexed by only one parameter and the error distribution normal. The model is as follows:

$$\begin{bmatrix} X_{i1} \\ X_{i2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \tau & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{bmatrix} \quad (i = 1, \ldots, n)$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independently $N_2(0, I_2)$.

The model is also equivalent to the following:

$$\begin{bmatrix} X_{i1} \\ X_{i2} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \tau \\ \tau & 1 + \tau^2 \end{bmatrix} \right) \quad (i = 1, \ldots, n),$$

where X_i $(i = 1, \ldots, n)$ are independent. The sufficient statistic is then $S = (\Sigma X_{i1} X_{i2}, \Sigma X_{i1}^2)$ and ΣX_{i1}^2 is obviously ancillary. One implication is that S is not complete which is in fact the case.

3.3 Equivariant estimation

We have seen that if we have a problem that remains invariant under G and such that $\nu(\theta)$ is a constant function, then the maximal invariant statistic $C(S)$ is ancillary. If we have an equivariant estimate $\hat{\theta}$ of θ and a loss function $L(\theta, \hat{\theta})$ that remains invariant under G^*, then, by a standard result, the risk $E_\theta\{L(\theta, \hat{\theta})\}$ does not

depend on θ. By a similar argument, we can show that the conditional

risk $E_\theta\{L(\theta,\hat{\theta})|C\}$ also does not depend on θ. Denote these by R and

R(C). Then $R = E\{R(C)\}$ and the fact that θ is not involved simplifies

interpretation. We have seen an application of this result in Example 1.

3.4 *An interpretation*

Within the context of transformation models, we can also give

an interpretation to the statement that two samples with the same

ancillary statistic value contain equal amounts of information. In

doing so, we make use of fiducial distributions which are distributions

on the parameter space. Buehler (1982) suggests using a more neutral

term "induced distribution". It seems inevitable that a fiducial or a

similar kind of argument is needed, because of the frequentist definition

of information as an average over the sample space or, in the case of

conditional information, part of the sample space. Thus it is difficult

to talk about information contained in a sample because we cannot condition

on the observed sample. As remarked by Efron (1978), "This is impossible

in the frequentist framework, since if we reduce our averaging set to

one data point, there is nothing left to average over." This is where

fiducial theory may help, because the method of fiducial probability

aims to get probability statements about parameters without the use of

Bayes' formula; probabilities concerning pivotal quantities are inverted

into formal statements about parameters. While this approach creates

much controversy, meaningful results are obtainable when applied to

transformation models. The interested reader is referred to Zacks

(1971, §§7.3-4).

The kind of interpretation that we are aiming at can be illustrated

by Example 3. Clearly $\hat{\theta} = (X_{(1)} + X_{(n)})/2$ is an equivariant estimate of

θ with respect to the translation group and $S = (\hat{\theta}, C)$ is minimal sufficient. The statement $\hat{\theta}|C = c \sim \theta + U(-(1-c)/2, (1-c)/2)$ can be inverted formally to $\theta \sim \hat{\theta} + U(-(1-c)/2, (1-c)/2)$, the induced distribution of θ. Thus if we have two samples $(\hat{\theta}, c)$ and $(\hat{\theta} + a, c)$, their corresponding induced distributions are simply translations of one another. As a result, whatever amount of support the sample $(\hat{\theta}, c)$ gives to any particular value of θ, say θ_0, the sample $(\hat{\theta} + a, c)$ will give the same amount of support to $\theta_0 + a$.

The general case is similar. We assume that the minimal sufficient $S = (\hat{\theta}, C)$, where $\hat{\theta}$ is an equivariant estimate and C is the ancillary statistic which in this case is also the maximal invariant statistic. If we have two samples $s_1 = (\hat{\theta}_1, c)$, $s_2 = (\hat{\theta}_2, c)$ with the same ancillary statistic value, $\hat{\theta}_2$ and $\hat{\theta}_1$ are necessarily related by $\hat{\theta}_2 = g^* \hat{\theta}_1$ for some $g^* \in G^*$. Let $f(\theta_1|s_1)$, $f(\theta_2|s_2)$ be the induced densities; then they are similarly related by g^*. To be precise, if θ_1 has density $f(\theta_1|s_1)$ and $\theta_2 = g^* \theta_1$, then θ_2 has density $f(\theta_2|s_2)$.

Most authors consider the maximum likelihood estimator as the information carrying part of the minimal sufficient statistic. In our discussion, any equivariant estimator can be used as the information carrier. Since the maximum likelihood estimator is equivariant, our treatment is more general.

4. PRECISION INDEX AND STATISTICAL CURVATURE

4.1 *Exact precision index*

In discussing the role of an ancillary statistic as a precision index, Buehler (1982) calls C an exact precision index if for some parametrization

$$f(\hat{\theta}|c;\theta) = f_0(\hat{\theta} - \theta|c), \tag{4.1}$$

where $\hat{\theta}$ is the maximum likelihood estimate. In other words, C is an exact precision index if conditionally, θ is a location parameter. An implication of this definition is that the unconditional and conditional Fisher information both do not depend on θ. This again simplifies interpretation. Example 2 is an illustration.

4.2 *Transformation models*

As remarked by Barndorff-Nielsen (1980), for transformation models, it is no essential restriction to assume that θ is conditionally a location parameter. It follows that, for transformation models, the ancillary statistic C is an exact precision index, and the statistical curvature γ_θ is constant. In passing, we mention that Amari (1982a) defines a one-parameter family of affine connexions with associated curvature. Efron's (1975) definition corresponds to the exponential connexion and is called by Amari the exponential curvature.

Let $\ell_\theta(X)$ be the log likelihood function of $X = (X_1,\ldots,X_n)$ and denote its first and second derivative with respect to θ by $\dot{\ell}_\theta(X)$ and $\ddot{\ell}_\theta(X)$. The total Fisher information is $I_\theta^{(n)} = E_\theta\{-\ddot{\ell}_\theta(X)\}$ whereas the conditional Fisher information $I_\theta(C)$ is defined as the Fisher information of the conditional distribution of X given C. Cox (1971) proposes the

use of var$\{I_\theta(C)\}$ as a measure of the extent to which the ancillary statistic C divides the possible data points into relatively informative ones and relatively uninformative ones. In our present case, $I_\theta^{(n)}$, $I_\theta(C)$, and γ_θ are all free of θ and we denote them by $I^{(n)}$, $I(C)$ and γ respectively. It can be shown that $\sqrt{n}\{I(C)/I^{(n)} - 1\} \longrightarrow N(0,\gamma^2)$. Without loss of generality we may assume $I^{(n)} = n$, so that under regularity conditions, var$\{I(C)\}/n \longrightarrow \gamma^2$ $(n \to \infty)$. Thus the statistical curvature is a measure of the effect of conditioning, in that the larger γ^2 is, the larger is the difference between conditional and unconditional inference. Our results can be summarized as follows.

THEOREM 1. *For transformation models,*

(i) the statistical curvature γ_θ does not depend on θ,

(ii) the ancillary statistic C is an exact precision index,

(iii) under regularity conditions

$$\sqrt{n}\{\frac{I(C)}{I^{(n)}} - 1\} \longrightarrow N(0,\gamma^2),$$

(iv) under regularity conditions

$$\lim_{n \to \infty} \frac{1}{n} \operatorname{var}\{I(C)\} = \gamma^2.$$

Proof. (iii) Without loss of generality, $I^{(n)} = n$. Efron & Hinkley (1978) prove that

$$\sqrt{n}\{\frac{-\ddot{\ell}_{\hat\theta}(X)}{n} - 1\} \longrightarrow N(0,\gamma^2),$$

where $\hat\theta$ is the maximum likelihood estimate of θ and $-\ddot{\ell}_{\hat\theta}(X)$ is the so-called observed information. To obtain our desired result, we need to show that

$$\frac{-\ddot{\ell}_\theta(X) - E_\theta\{-\ddot{\ell}_\theta(X)|C\}}{\sqrt{n}} \longrightarrow 0 \quad \text{in probability.}$$

To show this, we shall make use of the following two facts:

(a) $E_\theta\{-\ddot{\ell}_{\hat\theta}(X)|C\} = -\ddot{\ell}_\theta(X)$,

(b) $E_\theta\{n(\hat\theta - \theta)^2|C\} \to 0$ in probability,

the latter being proved by Efron and Hinkley (1978). By a Taylor's series expansion, we have $\ddot{\ell}_\theta(X) = \ddot{\ell}_{\hat\theta}(X) + (\theta - \hat\theta)\dddot{\ell}_{\theta_1}(X)$ where $|\theta_1 - \hat\theta| < |\theta - \hat\theta|$. Assume that $|\dddot{\ell}_\theta(X_1)| < M(X_1)$ for θ_1 in a neighbourhood of θ, and with $E_\theta\{M(X_1)\} = \mu < \infty$. Then, with probability 1,

$$\left|\frac{-\ddot{\ell}_{\hat\theta}(X) - E_\theta\{-\ddot{\ell}_\theta(X)|C\}}{\sqrt{n}}\right| \le E_\theta\{\sqrt{n}|\theta - \hat\theta|\sum\frac{M(X_i)}{n}|C\}$$

$$= E_\theta\{\sqrt{n}|\theta - \hat\theta|\mu|C\} + E_\theta\{\sqrt{n}|\theta - \hat\theta|(\sum\frac{M(X_i)}{n} - \mu)|C\}.$$

The proof is completed by applying the Cauchy-Schwartz inequality to the last two expectations and then using (b).

(iv) The conditions are those that guarantee the equality of the asymptotic variance with the limit of the variances. A set of such conditions can be found in Zacks (1971, p.244). In our case, the conditions can be simplified somewhat because $E\{I(C)\} = I^{(n)}$ so that the bias is zero.

4.3 Exponential models

Another general class of models for which exact ancillary statistics exist is the exponential class with a cut (Barndorff-Nielsen, 1978, 1980).

Consider the full exponential model of order k

$$p(x;\lambda) = \exp\{\lambda\cdot t(x) - \kappa(\lambda) - h(x)\}.$$

Here λ and t are both vectors of dimension k and $\lambda\cdot t$ denotes their dot product. If $\lambda = \lambda(\theta)$ where θ is d-dimensional, then we have a d-dimensional

curved subfamily of the full model. If we let $\tau(\lambda) = E_\lambda(t)$ and partition both t and τ into two components, say $t = (u,\dot{v})$, $\tau = (\xi,\eta)$, such that u and $\xi = E(u)$ are $(k-d)$-dimensional while v and $\eta = E(v)$ are d-dimensional, then the constraint $\xi = \xi_0$ also defines a d-dimensional curved subfamily. In this case, we can take θ to be the second component of the similar partition of λ, that is, $\lambda = (\chi,\theta)$ with χ some function of θ. If χ is of the form $\chi = \alpha(\xi) + \beta(\theta)$ for some function α and β, then u is a cut and then $C = u(X)$ is exactly ancillary. A number of examples of cuts are given by Barndorff-Nielsen (1978, §10.2). If we have n independent observations $X_1, \ldots, X_n$, then $\Sigma t(X_i)$ is sufficient and $C = \Sigma u(X_i)$ is ancillary.

The following example, due to Barndorff-Nielsen (1980) and used by Buehler (1982) has an ancillary statistic which is not an exact precision index but is neverthless an approximate precision index.

Example 8. Let $(X_1,Y_1), \ldots, (X_n,Y_n)$ be independent observations from the correlated bivariate gamma density (3.1). If α, θ are restricted by $\alpha - \theta^{-1} = a$ for any $a > 0$, the bivariate density becomes

$$f(x,y;\theta) = a\theta I_0(2\sqrt{xy})\exp\{-ax - (x/\theta) - \theta y\}$$

and the marginal distribution of x is Exp(a). The minimal sufficient statistic is $S = (\Sigma X_i, \Sigma Y_i)$ and $C = \Sigma X_i \sim \text{Gamma}(n,a)$ is ancillary. Buehler (1982) investigates the case $n = 1$ and finds that the precision of estimation depends not only on C but also on θ and that this property remains true no matter how θ is transformed. Since, there is no parametrization such that the conditional Fisher information is free of the parameter, C is not an exact precision index. However, numerical calculations indicate that C is an approximate precision index.

The next theorem, which should be compared with Theorem 1, shows that this phenomenon holds true in general.

THEOREM 2. *For exponential families with a cut,*

(i)   *the statistical curvature* γ_θ *is not constant,*

(ii) $\mathrm{var}\{I_\theta(C)\} = n\gamma_\theta^2 I_\theta^2$ *where* $I_\theta = E_\theta\{-\ddot{\ell}_\theta(X_1)\}$,

(iii) *the ancillary statistic* $C = \Sigma u(X_i)$ *is not an exact precision index,*

(iv) $(1/n)I_\phi(C) = 1 + \epsilon_n$, $\epsilon_n \to 0$ *a.s., where* $\phi(\theta)$ *is the variance stabilizing transformation for* I_θ.

Proof. (i)   Let

$$M_\theta = \begin{bmatrix} E_\theta\dot{\ell}_\theta^2 & E_\theta\dot{\ell}_\theta\ddot{\ell}_\theta \\ E_\theta\ddot{\ell}_\theta\dot{\ell}_\theta & E_\theta\ddot{\ell}_\theta^2 - I_\theta^2 \end{bmatrix},$$

where $\dot{\ell}_\theta$, $\ddot{\ell}_\theta$ denote the first and second derivatives of $\ell_\theta(X_1)$; then $\gamma_\theta^2 = |M_\theta|/I_\theta^3$.   In our case

$$p(x;\theta) = \exp\{\chi(\theta)\cdot u + \theta v - \kappa(\lambda(\theta)) - h(x)\},$$

where $\lambda(\theta) = (\chi(\theta),\theta)$, so that

$$\dot{\ell}_\theta = \chi'(\theta)\cdot(u - \xi_0) + (v - \eta),$$

$$\ddot{\ell}_\theta \quad \chi''(\theta)\cdot(u - \xi_0) - \partial\eta/\partial\theta.$$

From Barndorff-Nielsen's (1980) result that $\dot{\ell}_\theta(X_1)$ is uncorrelated with $u(X_1)$, we have

$$E_\theta\dot{\ell}_\theta\ddot{\ell}_\theta = \mathrm{cov}(\dot{\ell}_\theta,\ddot{\ell}_\theta) = 0$$

and

$$\gamma_\theta^2 = \{I_\theta(E_\theta\ddot{\ell}_\theta^2 - I_\theta^2)\}/I_\theta^3 = \mathrm{var}(\ddot{\ell}_\theta)/I_\theta^2,$$

since $I_\theta = -E_\theta \ddot{\ell}_\theta$. Thus γ_θ^2 is not constant.

(ii) $\ddot{\ell}_\theta(X) = \Sigma \ddot{\ell}_\theta(X_i) = \chi''(\theta) \cdot (\Sigma u(X_i) - n\xi_0) - n\partial\eta/\partial\theta$ is a function of $C = \Sigma u(X_i)$ and so

$$I_\theta(C) = E_\theta \{-\ddot{\ell}_\theta(X)|C\} = -\ddot{\ell}_\theta(X)$$

and

$$\mathrm{var}\{I_\theta(C)\} = \mathrm{var}\{-\ddot{\ell}_\theta(X)\} = n \, \mathrm{var}\{-\ddot{\ell}_\theta(X_1)\}$$

$$= n\gamma_\theta^2 I_\theta^2.$$

(iii) From (ii), $\gamma_\theta^2 = \mathrm{var}\{I_\theta(C)\}/nI_\theta^2$. Since the statistical curvature is invariant under reparametrization, there exists no parametrization such that I_θ and $I_\theta(C)$ are free of θ. In other words, C is not an exact precision index.

(iv) Let $\phi(\theta)$ be the variance stabilizing transformation for I_θ, so that $I_\phi^{(n)} = n$. Since $I_\theta(C) = -\ddot{\ell}_\theta(X) = -\Sigma \ddot{\ell}_\theta(X_i)$,

$$I_\phi(C) = -\Sigma \ddot{\ell}_\theta(X_i)/I_\theta$$

and

$$\frac{1}{n}I_\phi(C) = \{-\frac{1}{n}\Sigma \ddot{\ell}_\theta(X_i)\}/I_\theta$$

$$= \{I_\theta + \epsilon_n\}/I_\theta,$$

$$= 1 + \epsilon_n, \quad \text{where } \epsilon_n \to 0 \text{ a.s.}$$

In this sense, C can be regarded as an approximate precision index.

4.4 *Discussion*

In his discussion of (Efron, 1975), Cox (1975) comments that "Existence of an approximate ancillary must be connected with the approximate constancy of γ_θ as a function of θ; it would be good to

have the connexions explored." Cox's comment seems to suggest that the existence of an ancillary statistic is connected with the constancy of statistical curvature. This cannot be correct since an exact ancillary statistic exists *for exponential families with a cut.* Our results seems to indicate that the constancy of statistical curvature is related to the concept of exact precision index. Perhaps the clearest indication of this is in the proof of part (iii) of Theorem 2 where the non-constancy of γ_θ leads us to conclude that C cannot be an exact precision index.

5. MISCELLANEOUS REMARKS

As in most papers on this topic, we deal with scalar θ; this is for the sake of simplicity and also because we believe that it is often the simple examples which shed light on a concept. In principle, we can also deal with the multiparameter case. An example of a multiparameter transformation model is the hyperboloid distribution with fixed value of the "concentration parameter". That this is the case is not obvious; Barndorff-Nielsen (1980) attributes the proof to J.L. Jensen. On the other hand, an example of a two-parameter exponential family for which an exact ancillary exists is due to G.W. Cobb, cited by Hinkley (1980). The generalization of the concept of exact precision index is not as obvious, because covariance stabilizing transformations do not exist in general (Holland, 1973).

We also do not deal with nuisance parameters. For problems invariant under a group of transformations G, if $\nu(\theta)$ is not a constant function, then $\nu(\theta)$ becomes a nuisance parameter. Examples are model II of analysis of variance and the problem of estimating the common mean of two normal distributions $N(\mu, \sigma_1^2)$, $N(\mu, \sigma_2^2)$ based on two samples of equal size. For a treatment of equivariant estimation in the presence of a nuisance parameter, the reader is referred to Zacks (1971, §7.7) where Bayes equivariant and fiducial estimators are derived for the examples just mentioned.

REFERENCES

AMARI, S. (1982a). Differential geometry of curved exponential families. *Ann. Statist.* 10, 357-85.

AMARI, S. (1982b). Geometrical theory of asymptotic ancillarity and conditional inference. *Biometrika* 69, 1-17.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory.* New York : Wiley.

BARNDORFF-NIELSEN, O. (1980). Conditionality resolutions. *Biometrika* 67, 293-310.

BASU, D. (1955, 1958). On statistics independent of a complete sufficient statistic. *Sankhyā* 15, 377-80 and 20, 223-6.

BECKER, N. & GORDON, I. (1983). On Cox's criterion for discriminating between alternative ancillary statistics. *International Statistical Review* 51, 89-92.

BUEHLER, R.J. (1982). Some ancillary statistics and their properties. *J. Am. Statist. Assoc.* 77, 581-9.

COX, D.R. (1971). The choice between alternative ancillary statistics. *J. R. Statist. Soc.* B 33, 251-5.

COX, D.R. (1975). Contribution to discussion of paper by B. Efron. *Ann. Statist.* 3, 1221.

COX, D.R. (1980). Local ancillarity. *Biometrika* 67, 286-97.

COX, D.R. & HINKLEY, D.V. (1974). *Theoretical Statistics.* London : Chapman and Hall.

EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.* 3, 1189-242.

EFRON, B. (1978). Controversies in the foundations of statistics. *American Mathematical Monthly* 85, 231-46.

EFRON, B & HINKLEY, D.V. (1978). Accessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information (with discussion). *Biometrika* 65, 457-87.

FISHER, R.A. (1935). The logic of inductive inference. *J. R. Statist. Soc.* 98, 39-54.

FRASER, D.A.S. (1968). *The Structure of Inference.* New York : Wiley.

HINKLEY, D.V. (1977). Conditional inference about a normal mean with known coefficient of variation. *Biometrika* 64, 105-8.

HINKLEY, D.V. (1980). Likelihood as approximate pivotal distribution. *Biometrika* 67, 287-92.

HOLLAND, P.W. (1973) Covariance stabilizing transformations. *Ann. Statist.* 1, 84-92.

LEHMANN, E.L. (1981). An interpretation of completeness and Basu's theorem. *J. Am. Statist. Assoc.* 76, 335-40.

SPROTT, D.A. (1961). An example of an ancillary statistic and the combination of two samples by Bayes' theorem. *Ann. Math. Statist.* 32, 616-8.

ZACKS, S. (1971). *The Theory of Statistical Inference.* New York : Wiley.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
363	AD - A1600 30	

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
The Indexing Properties Of An Ancillary Statistic	TECHNICAL REPORT
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Anthony Y.C. Kuk	N00014-76-C-0475

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Department of Statistics Stanford University Stanford, CA 94305	NR-042-267

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Office of Naval Research Statistics & Probability Program Code 411SP	August 28, 1985
	13. NUMBER OF PAGES
	21

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ancillarity; Conditional Fisher information; Equivariant estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The role of an ancillary statistic as an information and as a precision index is discussed. For transformation models, it can be shown that, in some sense, two samples with the same ancillary statistic value contain the same amount of information. For both transformation and exponential models, the variance of the conditional Fisher information is proportional to the square of the statistical curvature. It appears that the constancy of statistical curvature is related to the concept of exact precision index.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73
S/N 0102-014-6601 |

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

11-85

DTIC